Systematic review of decision analytic modelling in economic evaluations of low back pain and sciatica

Hall, James; Konstantinou, Kika; Lewis, Martyn; Oppong, Raymond; Ogollah, Reuben; Jowett, Sue

DOI:
10.1007/s40258-019-00471-w

SYSTEMATIC REVIEW OF DECISION ANALYTIC MODELLING IN ECONOMIC EVALUATIONS OF LOW BACK PAIN AND SCIATICA

Running Title: Decision Modelling in LBP and Sciatica

Authors:

James A Hall[1]* MSc, Kika Konstantinou[1,2] PhD, Martyn Lewis[1,3] PhD, Raymond Oppong[4] PhD, Reuben Ogollah[5] PhD, Sue Jowett[1,4] PhD

Affiliations

[1]Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG, UK

[2] Haywood Hospital, Midlands Partnership Foundation Trust, Staffordshire, ST6 7AG, UK.

[3] Keele Clinical Trials Unit, Keele University

[4] Health Economics Unit, Institute of Applied Health Research, University of Birmingham, Birmingham, B15 2TT

[5] Nottingham Clinical Trials Unit, School of Medicine, University of Nottingham, Nottingham, NG7 2UH, UK

[*]Corresponding author

j.a.hall@keele.ac.uk

Telephone: 07450859296

ORCID: 0000-0001-8024-5427

Compliance with Ethical Standards: This research was undertaken as part of a PhD project, with a three-year research stipend for JAH funded by The Institute for Primary Care & Health Sciences, Keele University. KK was supported through a HEFCE Senior Clinical Lecturer award. ML, RaO, ReO, and SJ have no conflict of interest.

Key Points for Decision Makers:

- Despite methodological and technological advances in health economic modelling, the quality of modelling studies for low back pain and sciatica is generally poor.
- High quality modelling studies in both conditions, which reflect modelling guidelines, as well as evolving understanding of both conditions, are required to enhance the quality of economic evidence for treatments in both low back pain and sciatica.

Acknowledgements: Nadia Corp, Research Associate: Systematic Reviews at Keele University, assisted with development of the search strategy.

Word Count: 7384

Number of Pages: 27

Number of Figures: 1

Number of Tables: 5

Keywords: low back pain, back pain, sciatica, economic evaluation, decision model, decision modelling, systematic review

SYSTEMATIC REVIEW OF DECISION ANALYTIC MODELLING IN ECONOMIC EVALUATIONS OF LOW BACK PAIN AND SCIATICA

Abstract

Background:

Low back pain (LBP) and sciatica place significant burden upon individuals and healthcare systems, with societal costs alone likely to be in excess of £15bn. Two recent systematic reviews for LBP and sciatica identified a shortage of modelling studies in both conditions.

Objectives:

1. Document existing model-based economic evaluations for treatment and management of both conditions.
2. Critically appraise current modelling techniques, analytical methods, data inputs, and structure, using narrative synthesis.
3. Identify unresolved methodological problems and gaps in literature.

Methods:

A systematic literature review was conducted whereby 6512 records were extracted from eleven databases with no date limits imposed. Studies were abstracted according to a pre-designed protocol, whereby they must be economic evaluations which employed an economic decision model and considered any management approach for LBP and sciatica. Study abstraction was performed initially by one reviewer who removed duplicates and screened titles to remove irrelevant studies. 133 potential studies for inclusion were then screened independently by other reviewers. Consensus was reached between reviewers regarding final inclusion.

Results:

21 publications of 20 unique models were included in the review, of which five were modelling studies in LBP and 16 in sciatica. Results revealed a poor standard of modelling in both conditions, particularly regarding modelling

techniques, analytical methods, and data quality. Specific issues relate to inappropriate representation of both conditions in terms of health states, insufficient time horizons, and use of inappropriate utility values.

Conclusion:

High quality modelling studies, which reflect modelling best practice, as well as contemporary clinical understandings of both conditions, are required to enhance the economic evidence for treatments for both conditions.

1. Introduction

Low back pain (LBP) is a very common symptom experienced by most people at some point in their lives [1]. The lifetime prevalence and incidence of LBP are 80–85% [2] and 58-84% [53], respectively. Where back pain is long-lasting, and continues for more than three months, it is considered chronic. The prevalence of chronic LBP is 11% in men and 16% in women [4]. Most of LBP is termed non-specific as rarely a cause can be identified to explain the symptoms [1]. National and international clinical practice guidelines for the management of LBP recommend conservative interventions such as exercises, pain medication as well as psychological therapies, in addition to appropriate advice and patient education about the condition [5].

Sciatica (also called lumbar spinal radiculopathy or nerve root pain) is one distinct presentation of LBP, characterised by pain radiating to the leg and often into the foot and toes. The most common reasons for sciatica are compression or irritation of a lumbar spinal nerve root by a prolapsed disc, or tightening of the spinal or lateral canal, causing the condition termed spinal stenosis which may also give sciatic type of symptoms [6] [7]. Whilst the prevalence of sciatica is much less than that of LBP alone [8], it is considered responsible for much of the indirect costs and lost workdays associated with back pain [9] [10]. Whilst the management of non-specific LBP is almost exclusively conservative, sciatica management options can potentially include, spinal epidural injections and spinal surgery, for patients not improving sufficiently with conservative management [11]. Given the scale of the burden of disease, high-quality economic evidence is urgently required to determine cost-effective means of managing both conditions.

Generally, compared with a single within-trial-based analysis, decision analytic models are considered a more appropriate tool for conducting economic evaluations of chronic conditions [12]. Two previous systematic reviews [11] [13] identified mixed results regarding the cost-effectiveness evidence for treatments for both conditions, as well as a shortage of modelled cost-effectiveness analyses for non-invasive and non-pharmacological treatments for LBP [14] [15] and treatments commonly available for sciatica in the UK [16]. This study therefore aims at identifying and critically evaluating current modelling techniques in both conditions via narrative synthesis, with the aim of probing methodological issues and gaps in literature.

2. Methods

2.1 Study identification and selection

A protocol was developed using the PRISMA-P checklist (http://www.prisma-statement.org/Extensions/Protocols.aspx). The protocol specified that studies would be included if undertaking cost-effectiveness, cost-consequence, cost-benefit, or cost-utility analyses. Studies could consider any treatment or management approach for patients with LBP or sciatica. Reviews would be included if the publication also contained an economic model. This review excluded economic evaluations which did not include decision analytic modelling, studies which did not fully report methods, in this case conference proceedings and abstracts, and non-English language studies.

Articles were identified using database searches, with studies subsequently identified by reference searching also considered. The following databases were searched: MEDLINE, EMBASE, PsychINFO, CINAHL, AMED, EconLit, Cochrane DARE and CDSR, HTA, NHS EED, and Web of Science.

In developing the search strategy, JH produced an initial draft strategy where economic terms were based upon a strategy developed by the National Health Service (NHS) Centre for Reviews and Dissemination at the University of York. Clinical terms reflected strategies taken by other systematic reviews of economic evaluations in LBP and sciatica [11] [13]. Terms used in papers known a priori to require inclusion were also used. The strategy was subsequently refined by the co-authors, a copy of the search strategy for MEDLINE is included in online appendix 1. Database searching took place during January 2017. The review was subsequently updated in February 2019.

2.2 Data selection and extraction

Figure 1 details the selection process. 6512 records were imported into Endnote, of which 1556 were duplicates. A two-stage exclusion process was employed. Of the 4956 unique studies, 4823 were excluded in accordance with the protocol exclusion criteria, by first reviewer (JH). 3762 studies were excluded because they did not reflect the clinical area, 877 were not economic evaluations with economic model, 175 were abstracts, and 9 conference publication. 10% (or 400) of these excluded studies were independently checked by one other reviewer (SJ). The 133 titles that were considered potentially relevant were then subdivided into “included” and “possible” by first reviewer (JH). A second reviewer (SJ) checked the suitability of “included” studies, whilst all “possible”’ studies were independently reviewed by all four other reviewers (SJ, KK, RaO and ReO) for relevance. In Figure 1, publications deemed “clinically irrelevant” were studies that the first two authors (JH) and (SJ) were not able to determine were sciatica or non-specific LBP, these were excluded following further consultation with the clinician on the team (KK) regarding whether or not the various conditions described in the studies, constituted a diagnosis of sciatica or non-specific LBP. Having reviewed all of the possible studies independently, the five authors (SJ, KK, RaO, ReO, and JH) then reached consensus regarding inclusion of the final 20 studies, and one additional study (which was referenced in one of the original 20) was also added.

Data was extracted in accordance with pre-specified criteria detailed in the study protocol.

3. Results

3.1 Overview of studies

Table 1 provides an overview of studies included in this review. Studies were classified according to whether patients had LBP or sciatica, as each condition potentially requires a different modelling approach. We further subdivided sciatica studies according to whether they contained a non-surgical comparator, as modelling studies solely for surgical treatments should require a unique structure with model time horizon beginning at the point the patient undergoes surgery.

Five studies were included that evaluated the cost-effectiveness of treatments or management of LBP. Of the fifteen that focused on sciatica, nine studies (eight unique models) considered at least one non-surgical treatment, and seven described solely surgical treatments for sciatica. A variety of interventions were considered, with analyses conducted across various perspectives. Nearly all included studies were published after 2010 (n = 18). Half were conducted in the U.S (n = 11), and the remainder were mostly from developed countries (n=9).

3.2 Model Design and Structure

3.2.1 Model type

Table 2 summarises the characteristics of model design and structure. Most studies with a determinable structure (n = 19) used a Markov model (n = 13) or decision tree (n=4), with two using a decision tree prior to their Markov model.

3.2.2 Time horizon, cycle length and discounting

A variety of time horizons were employed in both conditions.

For LBP models, horizons were as short as one episode of pain [17] to as long as lifetime [18] [19]. Time horizon was associated with model type, with the decision trees modelling the short-term horizons [17] [15] and Markov models employed for the longer-term horizons [14] [18] [19]. In terms of cycle lengths, the three Markov models which stated their cycle length, all used a cycle length of three months [14] [18] [19]. All studies over year in duration ought to discount according to national guidelines [35], and all LBP studies did so appropriately.

For the sciatica models which included a non-surgical comparator, models were as short as twelve months [20] [11] [21] and long as ten years [50]. Markov models were employed to perform the analyses in the longer-term horizons [16] [25] [50]. Three months was the most common choice of cycle length, with two models employing such a length [16] [24]. All models with time horizon over 12-months discounted appropriately, aside from one which explicitly ruled out discounting, incorrectly [23].

The sciatica models which included only surgical options generally modelled longer time horizons, with the shortest being two years [31], and longest models including horizons of ten years [26] [27]. However, many of the cycle lengths used in these models were unclear. Of those models using time horizon over 12-months, one explicitly ruled out discounting, incorrectly, on the basis their time horizon was 2 years [31]. In three studies it was unclear whether discounting was included, or what the time horizon to determine the need to discount [32] [29] [30].

3.2.3 Health states

Two LBP models based their model structure upon treatment success [15] [17], whilst Kim et al. [14] used states reflecting temporal classification of LBP, e.g based upon episode duration. Both Wielage et al. [18] [19] publications used states consistent with the model’s interest in pharmacological treatment. Six of the non-surgical sciatica treatment models with identifiable structure used “treatment success” to structure their models, all models also included symptom recurrence or second treatment or reoperation. Tapp et al. [50] used a simple four state alive (post-surgery)/dead model, with the addition of complication and surgical states. Fitzsimmons et al. [21] is an exact replica of a model in a monograph by Lewis et al. [11], which attempted to replicate the treatment pathway for sciatica patients in the UK by using a triple stepped treatment pathway. Igarashi et al. [20] was the only model in the review which explicitly used health states directly reflective of pain severity, using mild or no pain, moderate, and severe pain.

Many of the surgical sciatica treatment models did not clearly report model structure and methodology [26] [28] [32] and one [29] used a simplistic structure. Of the remaining three state transition models, two [30] [31] structured their Markov model around treatment success. Meanwhile the simple four-state structure used by Kim et al. [27] enabled the representation of many events a surgical patient may experience, including the possibility that a patient may not choose a re-operation despite worsening of symptoms.

3.3 Parameter sources and methods of derivation

Whilst the origins and methods of derivation for treatment efficacies and recurrence rates were of reasonable quality, and therefore are not discussed further, there were particular concerns over the derivation of utility values, as well as resource use estimates.

3.3.1 Parameter sources for utility values

Table 3 summarises the sources, as well as the methods, used to calculate model parameters. Wielage et al. [18] [19] obtained their utility scores from meta-analysis, Norton et al. [15] and Kim et al. [14] both used utility values from adequately sized samples. However, the utility values for both the non-surgical treatment models for sciatica [11] [21] [23] [22] [24] and the solely surgical models [28] [31], were commonly derived from small trials with below 150 participants in total.

Both classes of sciatica models had problems with the use of utility values. Launois et al. [16] do not report anything about the origin of their values other than they were derived from a “patient survey”. Three studies in the review made use of the utility values for “severe LBP” patients in the Beaver Dam Health Outcomes study [34]. One such study Koenig et al. [25] state they used utility values for patients with herniated intervertebral disc from a previous economic evaluation [33] of lumbar discectomy for treatment of herniated intervertebral disc. However, in their original analysis [33] the authors stated they had used utility values for “severe LBP” from the Beaver Dam study [34], however Koenig did not state they were for severe LBP.

Two of the surgical treatment models [30] [26] using the utility values for “severe LBP” patients from the Beaver Dam study, used them for patients with lumbar spinal stenosis and spondylolisthesis. Bydon et al. [30] failed to acknowledge that the utility values were not for their target population. Moreover, as is explained in section 3.3.2, the latter two studies [26] [30] use unusually high utility values for symptom resolution. This is because they have used the utility values for individuals in perfect health in the Beaver Dam study [34].

Furthermore, the utility values used by Lewis et al. [11] and Fitzsimmons et al. [21] for “successful treatment” is actually the utility value achieved by early surgery in the van den Houdt et al. [47] study, which both papers [11]

[21] then apply for successes of all interventions in the model. Moreover, the utility value for “treatment failure” used by both [11] [21] represents the baseline utility in one randomised controlled trial (RCT).

Another study made an erroneous calculation [22] - reporting a 24-month QALY gain of 0.252 for ESI in their Table 3, comprised supposedly of 12 2-monthly QALY gains of 0.021, this is difficult to spot because they write (12 x 0.21) instead of what they presumably mean (12 x 0.021). However the calculation is incorrect aside from type error, the original source for the 0.021 QALY gain, clearly state, the gain is over 3 months [48].

3.3.2 Utility Values

Table 4 presents the actual utility values used as inputs in the studies in this review. There were some differences in the utility values used in the LBP studies, likely partly explained by their specific population. The meta-analyses of chronic LBP patients who were prescribed pharmacological treatments, showed that chronic LBP patients had utility values of between 0.7282 (Pregabalin) and 0.7688 (Naproxen) [18] [19]. Meanwhile, the other studies suggested lower values; Kim et al. [14] used 0.62 and 0.65 for chronic LBP patients on usual care and acupuncture, respectively, although acute LBP (0.85) and “well” states (0.96) were considerably higher; and Norton et al. [15] used much lower scores, with an improving chronic LBP patient having utility of 0.640, and a non-improver having utility of 0.59.

There was some consistency across some of the sciatica decision models which incorporated conservative care and used utility values independent of treatment. Igarashi et al. [20] report that a sciatica patient without pain would have a utility of 0.867, whilst “severe pain” would be 0.611, somewhat consistent with the scores used by Koenig et al. [25] for treatment of herniated intervertebral disc. Both studies are consistent also with the value used for “improvement” in sciatica patients by Lewis et al. [11] & Fitzsimmons et al. [21], although the “non-improvement” score in the latter is lower than other scores at 0.37.

Skidmore et al. [23] use utility scores for each treatment which are weighted averages of both improvers and non-improvers on each treatment, and are weighted also for likelihood and disutility of adverse events. Their weighted average for conservative care (which has only a 4.8% success rate), is a utility value of between 0.61 and 0.65 over the duration of the model. Given that non-improvers are the predominant constituent of the conservative care group, this value is consistent with the utilities of the other non-improving patients in the prior four studies discussed. The values used by Tapp et al. [50] for post –surgery, which represents a patient after surgery without complications, is slightly higher at 0.77, although the values also account for the disutility associated with complications and recurrence separately.

For the solely surgical studies, the two studies which used the Beaver Dam Study, applied a utility value of 0.79, for symptomatic spinal stenosis patients in one study [26] and for lumbar spondylolisthesis patients with a negative outcome in another [30]. Both studies also used a utility value of 0.97 for having a positive outcome. These values are far higher than the two other studies to use ”improve” / “non-improve” to differentiate patients. For patients with with lumbar spondylolisthesis; Kim et al. [27] used 0.74 for an improvement with fusion and no-fusion, compared to non-improvement of 0.50 and 0.54 respectively, baseline patients had a utility value of 0.58 [27]; and Schmier et al. [28] used a value for “clinical success” of 0.692, whilst “failure” had utility 0.552 and “worsening pain” 0.599.

Parkinson et al. [31] used much lower baseline pre-operation utilities of 0.42 and 0.36, although possibly explained by their specific population, patients with sciatica who had failed conservative treatment. After two years their patients who had AIDR had average utility of 0.67 and patients who had fusion had utility of 0.69. These scores are similar to the patients undergoing spinal fusion to whom Vertuani [32] assigned utilities of 0.72 after two years following minimally invasive surgery and 0.68 following open surgery.

3.3.***3*** Parameter sources for resource use

A variety of methods were used in deriving resource use.

Of the LBP studies, two studies [14] [15] estimated their resource use using information from trials, both of which were pragmatic trials. Lloyd et al. [17] used expert opinion to estimate resource use across the 4-day model time horizon. The two studies by Wielage et al. [18] [19] used a combination of expert opinion, databses and published literature. Of the non-surgical sciatica models, heavy reliance was made upon experts in order to derive resource use [11] [20] [23] [21]. Launois et al. [16] obtained resource use from the uncited survey and unreferenced material. Koenig et al. [25] relied primarily upon the SPORT trial [70], with surgery frequency derived from a recent analysis of the Medicare claims database [71]. Parker et al. [24] used data from a telephone interview to obtain follow-up costs. Tapp et al. [50]did manage to utilised the Medicare Provider Analysis and Review database for surgery, complication and re-operation costs, although the value of their analysis was diminished somewhat by their assumption that the cost of conservative care, as well as supplementary care for surgical patients, was zero. The solely surgical modelling publications made more use of administrative databases, with two studies deriving resource use almost entirely from hospital databases [26] [27], and two [28] [31] using databases in conjunction with published literature and expert opinion. Bydon et al. [30] used the same 137 patient institutional series they used to derive all their other parameters, which allowed the precise calculation of costs of surgery and inpatient costs at their institution. Vertuani et al. [32] performed a literature review and meta-analysis to identify resource usage. Finally, Yaghoubi et al. [29] did not identify resource use, but instead opted to obtain costs directly from published literature supplemented by patient feedback.

3.3.4 Methods and parameter sources for calculating societal costs

Six models attempted some analysis from the societal perspective. The most detailed attempt came from Igarashi et al. [20] who based their analysis on the Work Productivity and Activity Impairment (WPAI) questionnaire adapted for LBP, and includes losses owing to absenteeism and presenteeism. Three studies employed variants of the human capital approach to estimate productivity losses occurring due to productivity losses [14] [11] [24]. One study [23] claimed that “Indirect costs of lost productivity and intangible costs of pain and suffering related to treatment morbidity are not estimated directly but instead are implicitly incorporated in the utility values”

3.4 Summary of cost-effectiveness results

The results can be summarised as follows. For episodes of acute LBP, heat wrap dominates paracetamol and ibuprofen [17]. In Korea, acupuncture appears to be a cost-effective means of managing chronic LBP, relative to routine care [14]. Duloxetine is suggested to be a cost-effective means of managing chronic LBP pain in the U.S and Canada, and dominates most other pharmacological comparators, although it is unclear whether or not it is cost-effective relative to Naproxen (although there are further author correspondences regarding the modelling of tapentadol) [18] [19]. CBT (Cognitive Behavioural Therapy) is cost-effective relative to educational materials in treating chronic LBP in Canada [15]. Societal analyses considerably decreased the cost per QALY for the most effective treatments (i.e. made an intervention more cost-effective) [14] [19].

In sciatica, a comparison of two pharmacological approaches for managing Japanese patients who had severe LBP pain alongside a neuropathic component, showed that pregabalin was cost-effective compared to usual care treatment with standard analgesia [20]. All other sciatica studies considered some form of surgery as a comparator, with the length of time that patients spend receiving conservative care, and ordering of treatments, seeming to determine the cost-effectiveness of surgery. For example, in a review of over 100 different potential treatment combinations for patients presenting with sciatica, Lewis et al. [11] and Fitzsimmons et al. [21] show that in the UK, stepped care approaches based on initial treatment with non-opioids are the most cost-effective, whilst referring patients who fail an initial treatment to surgery is unlikely to be cost-effective. However, evidence from the U.S suggests that surgery after one course of failed treatment or extended duration of symptoms could be cost-effective. In one study of patients with moderate to severe lumbar spinal stenosis (LSS), who failed conservative therapy, minimally invasive lumbar decompression maybe cost-effective ($43, 760 per QALY) relative to standard non-surgical treatment [22]. Compared with ongoing conservative care, Skidmore et al. [23] show that for moderately impaired patients with LSS, decompression surgery using a spacer is cost-effective relative to non-surgical care and dominates laminectomy. Amongst LSS patients complete six months of conservative treatment without improvement, Parker et al. [24] show that the minimally invasive interspinous spacer and decompression surgery are highly cost-effective relative to conservative care at $16,300 per QALY and $15,200 per QALY respectively. Tapp et al. [50] show that over ten years, for patients with LSS and no previous surgery, minimally invasive procedures using a spacer and decompression are cost-effective relative to usual care at $25,000 per QALY and $30,874 per QALY respectively. Although the evidence on surgery for patients with lumbar disc herniation suffering functional limitations, suggests discectomy may not be cost-effective relative to non-surgical care ($52,416 per QALY), although inclusion of societal costs decreases the incremental cost effectiveness ratio (ICER) to $35,146 per QALY [25]

Of the papers which evaluate solely surgical techniques, three U.S studies suggest that various spinal fusion techniques are not cost-effective across different populations and relative to other surgical techniques without fusion, e.g. laminectomy without fusion [26], lumbar decompression without fusion [27] and the Coflex® interlaminar stabilization following decompressive laminotomy [28]. Furthermore, evidence from Iran supports the cost-effectiveness of the Coflex® relative to the X-Stop and laminectomy [29]. In studies evaluating only means of performing fusions, for US patients with degenerative sponydylolisthesis, interbody fusions were found cost-effective relative to non-interbody fusion, at $9,883.97 per QALY [30]. Amongst Australian patients with radicular pain who failed conservative treatment, posterolateral fusion was deemed to be the most cost-effective surgical approach for lumbar fusion or artificial intervertebral disc replacement (AIDR) [31]. Finally, in the UK & Italy, minimally invasive surgery dominated open surgery for one- or two-level lumbar spinal fusion in the treatment of degenerative lumbar spinal conditions [32].

Discussion

4.1 Statement of principal findings

4.1.1 LBP models

Structure

Choice of time horizon should be justified, and ought to reflect important differences between comparators and not be driven by data availability [35] [36] [37] [38]. The two LBP models by Wielage et al. [18] [19] were the only ones to use the life-time horizon, possibly because data was readily available for their specific decision problem. For the other three studies, the use of medium-term time horizons could be permissible given that many treatments for LBP do not provide a long-term treatment effect.

In terms of model type, four LBP models used Markov state transitions suggesting that these authors considered this approach appropriate for modelling LBP. However, in situations in which an early event or patient characteristic determines future patient pathways, and the number of health states required might be unwieldy, individual level microsimulation methods should be considered[36].

Health states should adequately reflect the condition specific health processes [36]. Upon this basis, the three-state approach used by Norton et al. [15], ”improved”, ”not improved” and ”dead”, might be considered an oversimplification. Whilst using “improved” and “not-improved” will, to some degree, always incorporate the pain associated with each treatment; unless utility values are always collected separately for each treatment, interventions which deliver higher rates of long-term improvement at a lower level of utility will be advantaged by using states reflecting “improvement”.

For LBP, we advocate an approach starting from the structure used by Kim et al. [14], who chose states reflecting a temporal classification of LBP - ”acute LBP”, ”chronic LBP”, “well”, and “dead”. However, the internal validity of their particular model is limited by the structural assumption that one recurrence of back pain in the “well” state can move the patient into “chronic LBP”. This leads to the possibility a patient with only two episodes of LBP across 5 years would be considered to have “chronic LBP”. Nonetheless, the modelling of degrees of symptoms is preferred over a dichotomy, as it is likely to produce results which better reflect the patient experience. It is also worth noting that it is likely that there is heterogeneity in pain severity within temporal classifications.

Current research evidence suggests that a potentially more appropriate categorisation of patients with LBP could reflect pain severity as well as the rate of recurrence [39]. Dunn et al. [39] for example, show that pain level after one year is predictive of pain level at seven years, and patients are categorised in three groups, according to persistence and severity, “no or occasional pain”, “persistent mild pain”, and “persistent severe pain”, with a fourth category used for those who show no consistent pattern. Ultimately, model structure and health state selection should involve consultation with subject experts and stakeholders [36]. Whilst experts were clearly involved in the construction of these models, none of the studies clearly justified, or discussed issues related to their model structure. The lack of discussion around choice of health states, or model choice and time horizon, is problematic for improving the representation of both conditions in model form, as such subjective components of the modelling process should be predicated upon a clear understanding and subsequent critique of, the principles upon which such decisions are made.

Data

High-quality information on resource use in LBP patients does not seem to be available. Both the Wielage et al. [18] [19] papers had to rely mainly upon expert opinion in order to derive resource usage, with two other studies relying upon trial-based estimates [14] [15]. With regard to the utility values for the LBP models, these were obtained in accordance with best practice [14] [15] [18] [19].

Four LBP models extrapolated over a longer period than their data allowed, although only Wielage et al. [18] [19] attempted lifetime extrapolation. Having evidence available to them, both studies modelled across the lifetime by adjusting the adverse event (AE) profiles according to age-dependent relative risks and used age-dependent utility values directly obtained from available literature. However, often in decision problems relating to LBP, evidence of long-term treatment effect is likely to be unavailable because trials commonly span only across one year of patient observation. Norton et al. [15] and Kim et al. [14] show how an incomplete evidence base can still be extrapolated in a decision analytic model by using assumptions and expert opinion in addition to literature. Both studies defined their health states independently of treatment and established utility values for those specific health states, therefore requiring only information on the long-term movement of patients between health states. Lacking data on long-term treatment efficacy, Norton et al. [15] assumed a gradual loss in efficacy (resolution of symptoms) over time of 20% per annum. Similarly, Kim et al. [14] extrapolate short-term treatment efficacy assuming short-term relative risks between the treatment arms remained constant over time.

Given that parameter behaviour over time, such as the long-term treatment effect, often represents the largest source of uncertainty within a model [40], studies performing extrapolation should be expected to undertake rigorous examination of temporal uncertainty [35]. However, none of the five LBP models really addressed temporal uncertainty, despite the importance of extrapolation assumptions used by Norton et al. [15] and Kim et al. [14]. Consequently, the reader is left without an understanding of how the uncertainty over the long-term treatment effect could impact the cost-effectiveness of these treatments for LBP.

4.1.2 Sciatica non-surgical treatment models

Structure

For this group of models, three studies used Markov modelling and three decision trees. However, given that surgery was a comparator in all models, and would be expected to improve long-term outcomes for sciatica patients, short time horizons modelled within a decision tree may seem unsuitable in this condition. The use of individual sampling models (ISM) could be of real value in this condition, given that candidacy for surgery is likely to be event-dependent, e.g. having a failed previous treatment, and/or time spent in severe pain.

Treatment guidelines for sciatica in the UK follow a stepped pathway, in that patients can receive more invasive treatments dependent upon prior treatment failures. Most models did allow between one and three stepped treatment failures before they allow surgery to take place e.g. [11] [16] [21] [24] [25]. Lewis et al. [11] and Fitzsimmons et al. [21] presented not only the stepped nature of the treatment pathway but provided 100 different treatment combinations. Yet, the complexity of the representation of the treatment pathway appears in contradiction to the simplicity of the estimates used for the utility values, which in turn may limit the validity of the study results. Whilst there is clearly a need to represent the stepped pathway when modelling a pre-surgical period of the treatment pathway, it is doubtful that a typical model should consider all combinations of possible treatments unless that is the specific goal of the analysis, as was the case in the Lewis et al. [11] and Fitzsimmons et al. [21] publications.

If the aim of the analysis is to compare an intervention with usual care, then it might be more efficient to have a comparator which represents common/usual practice. Indeed, most models only compared 2 or 3 treatment options, one of which included conservative care. Two studies [23] [24] used a comparator which reflected a specific combination of usual care treatments based upon an observed combination of treatments. For example, Skidmore et al. [23] use “conservative care” as their comparator, defined as at least one epidural steroid injection, supplemented by nonsteroidal anti-inflammatory medication, oral steroids, analgesics, physical therapy, or spinal manipulation therapy. Their data on treatment was derived from a trial where patients received usual care as considered appropriate for the individual patient. Finally, Parker al. [24] refer to conservative care, as physical therapy, pain medications (NSAIDs, mild opioids), and epidural steroid injections, as guided by clinical judgment of the treating physician. Their data were derived from analysis of institutional registry data, and trial data.

The approach used by The Institute for Clinical and Economic Review [46] in it’s model of treatments for lumbar disc herniation, whilst grey literature and not included in the review, is nonetheless worthy of consideration. It uses a Markov model which allows patients to continue on some specific combination of usual care treatments (identified by systematic review), and also allows a specific proportion of the cohort to move into receive a discectomy. Upon receiving a surgical procedure, patients can then improve, receive a second reoperation or suffer a complication.

Assuming that the model requires the entire pathway to be modelled, we would advocate taking the best of these approaches, by representing the stepped nature of the treatment pathway by initially defining usual care as the treatment combination which evidence shows that sciatica patients initially receive. If the model took the form of an ISM or a discrete event simulation, it would be possible to use tracker variables to track the length of time a patient remained symptomatic, and/or receiving the described initial usual care. After some defined period of non-improvement, a patient could either become eligible for a more intense usual care treatment option (a second-step), and/or upon failure of that second treatment become eligible for surgery. As well as time spent in receipt of usual care, patient’s candidacy for surgery could also be a function of the unique characteristics of that patient, although the degree of complexity will be limited by available data.

With regards to the choice of the states, similarly to the LBP models, all but one [20] used states relating to treatment success. The use of health states such as "improved" and "not improved" are perhaps more appropriate in models for sciatica in which perhaps interventions such as spinal injections and/or surgery might be expected to provide a more pronounced and sustained treatment response, on average, when compared to treatment effects for non-specific LBP.. But, as noted above, with respect to modelling, if such states are used, this will necessitate that utility values are collected for each intervention, as some interventions could deliver their improvement at very different utility levels. A related issue is that all of the studies use different definitions of success or improvement, making comparison between studies difficult.

Guidance suggests that where states reflect the treatment pathway effect, this ought to be justified and alternative methods of doing so explored in sensitivity analysis [40], however none of the studies justified their approach or explored alternate methods. The only model in this review to provide health states based upon pain severity [20] provides a starting place for how pain severity could also be used to conceptualise health states for sciatica. No pain/mild pain, moderate pain, and severe pain are used as states. Nonetheless, the inability to move between states during 3-11 months and the model's short 12-month timeframe, somewhat diminishes its potential as an example of best practice, if it is accepted that the analytical framework ought to be over a longer-period. Whilst "improved" or "success" models are likely to be conceptually linked with pain severity to some degree, the potential implications for cost-effectiveness estimates of using an "improved" or "success" based model over one based upon pain states could be one area of future research.

Data

Only 3 of 9 models attempted extrapolation. Launois et al. [16] used literature to extend parameters over the longer-term, whilst Igarashi et al. [20] simply extrapolated 8 week pain scores to 52 weeks, citing a previous study as justification. Tapp et al. [50] extrapolated from 4 years to 10 years, by determining the long-term rate of re-operation for decompression from the literature, and lacking data on long-term reoperation with the spacer, assumed the rate was identical to decompression. They also use the same utility values over time, and assume that complication rates are the same regardless of first surgery or re-operation. Their cost-estimates are unlikely to be accurate however, given that they do not include costs of other treatments aside from surgery and conservative care and post-surgical care carry zero cost.

That only 3 of 9 non-surgical sciatica models attempted extrapolation, was typically justified by the claims of Fitzsimmons et al. [21] of a "lack of evidence regarding relapse and recurrence rates" making "it difficult to extend the analysis beyond this time period". It is certainly the case that where estimates extend well beyond known available data, the accuracy of the estimates may be questionable. Yet as stated above, best practice guidelines in decision analytic modelling state that data availability should not define the time horizon of the model [35]. The apprehension regarding extrapolation is perhaps somewhat unfounded given the two of the LBP studies above [14] [15], and one sciatica model [50] made simple assumptions regarding the long-term treatment effect. Moreover, there are many techniques to infer values for unobserved model parameters [40]. For example, Mahon [40] advocates fitting parametric functions through statistical methods, and/or using expert opinion to derive some probabilistic assessment of the likelihood a parameter function takes a certain shape. Nothing approaching this level of sophistication was attempted in any of these papers in this review.

Given the productivity costs associated with both LBP and particularly sciatica, it might be expected to see more analyses performed from the societal perspective, at least as a secondary evaluation. Yet as Table 1 showed, and Table 3a details, in this review only six models performed some form of societal analysis, which were of varying rigour. The most detailed method [20] used the established methodology of the WPAI scale [41], a validated method to assess lost productivity that assesses the losses due to both absenteeism and presenteeism. Although for a full societal analysis, other non-medical costs should be also included, as per Kim et al. [14] who included costs of visiting, treatment times and traffic expenses.

The utility values used in this class of models reflect to some degree the lack of availability of utility data for sciatica patients. The methods used by Lewis et al. [11] and Fitzsimmons et al. [21] to calculate utility values show why it is not necessarily advisable to have values based upon the success or failure of a treatment because an appropriate analysis would require many unique utility values. As this information is clearly not available, the authors used the same utility values for "success" or "failure" of all the treatments in their study, which would bias against those treatments which delivered higher utility gains where successful and/or treatments which minimise utility losses where they fail. Moreover, whilst their value of 0.83 for treatment improvement seems consistent with other studies [20] [25], their "non-improvement" score of 0.37 is significantly lower than values used for non-improvement in most other studies in this review. This highlights the need for consistent health state selection and definition, in the models. Admittedly, the authors of both studies [11] [21] acknowledged the problems with utility values which they attempted to address using alternate scenarios in sensitivity analyses, although results of these analyses were not presented.

There were also concerns in relation to the derivation of resource use estimates, which ought to originate from real clinical practice instead of clinical trials [42]. Yet, for the sciatica non-surgical treatments, only Tapp et al. [50] obtained some of their cost estimates from a database reflecting actual patient healthcare usage, and half had to rely on expert opinion, suggesting a problem with availability of high-quality resource use information. This is understandable given how challenging it can be to identify and attribute visits when conditions may be mentioned only incidentally.

4.1.3 Sciatica surgical treatment models

Structure

Whilst it is commendable that the time horizons considered by the surgical intervention models were longer than the non-surgical models, albeit this group of models was replete with methodological problems. Of the models with identifiable structure, 5 out of 7 were Markov models, which are recommended for decision problems, such as modelling surgical procedures, where model horizon is longer and/or model contains time varying transition probabilities [44]. In the case of models which begin from a surgical process, a Markov model is sufficient, provided they allow the possibility of future re-operation following recurrence of symptoms. The need for an ISM is less pressing than in the case of a sciatica model including a pre-operative period, although allowing a risk of recurrence to be dependent upon individual characteristics would offer the model more flexibility.

Given that all comparators in this group were surgical procedures and did not feature conservative care, health states generally are reflective of the “success” or “failure” of surgery rather than health specific processes. As noted above, there is still a need amongst these models to use consistent health states and definitions for comparability. The structure used by Kim et al. [27] could provide a basic template for developing a more sophisticated model structure. Their model health states were potentially more appropriate because of the use of states such as “well”, “unwell”, or “return of symptoms”, which partially reflect the condition specific health processes, and captured all important *events* for sciatica patients, e.g. relapse, reoperation (including the flexibility of choosing whether or not to re-operate), clinical worsening and improvement, and general and perioperative death.

Data

The Kim et al. [27] structure can also provide a framework within which to engage in extrapolation processes. Using the health states of “well”, “unwell”, or “return of symptoms”, they then derived long-term utility values associated with these states, independent of treatment approach, from their one-year observational study. To justify the use of these one-year values across ten years, they referenced a study by Weinstein et al. [43] who suggested that utility values derived at one year are relatively stable across four years, and then assumed further stability to ten years. Although the authors then adjusted the values downwards by 3% per-year to account for clinical deterioration rather than present net value. . Where possible, the impact of age upon the utility of the patient population should be modelled using available data or a data driven assumption. Regardless, having established their utility values, they then required only annual transition probabilities between states for 10 years, which they derived using an administrative database study and data from the senior surgical authors practice.

As with the sciatica non-surgical models, the quality of the utility values was generally poor, as would be expected given that they are for similar populations. However, whilst it is to be lauded that attempts were made to derive patient- and treatment-specific utilities, small sample sizes [28] [31] and the use of LBP utility values from the Beaver Dam study [26] [30] were a common limitation of this group of publications. The use of the Beaver Dam values of 0.97 for being symptom free [26] or having a positive outcome [30] and 0.79 for having CLBP are quite different to the utility values used in other studies.

The use of administrative databases in five studies [26] [27] [28] [30] [31] facilitates precise calculation of resources associated with each treatment, and therefore offers more confidence in accuracy of the costs associated with these surgical procedures.

4.2 Strengths and weaknesses of the review

This is the first systematic review to identify, document and classify model-based economic evaluations of treatments for LBP and sciatica. It is possible that the search criteria may prove restrictive, in that model-based economic evaluations faced exclusion if they did not contain both economic and modelling terms. The search strategy may have been improved by including the term “utility”. However, the use of a very broad search strategy was employed including “economic” or “model” as standalone terms, with the specific aim of increasing potential number of studies. The breadth of these search criteria, as well the variety of databases used, is a key strength of this review, and the addition of a single paper identified from reference lists demonstrates that the search was exhaustive.

4.3 Implications for researchers, clinicians and policymakers

This review identifies flaws, and suggest opportunities, in models evaluating interventions for LBP and sciatica. Concerns relating to studies in this review include, not only modelling across inadequate time horizons, but the inappropriate use of utility data, calculation errors, a lack of transparency regarding methodologies, and the failure to consider the extent to which uncertainty and assumptions limit the applicability and generalizability of the results.

Overall, the current cost-effectiveness evidence is indicative of the uncertainty around the clinical effectiveness evidence on these treatments to-date. Most of the studies included in this review do report on the limitations of available effectiveness data in order to populate models. Policymakers’ attention is directed to the sensitivity analyses in these studies, which in some cases help with accounting for the uncertainty of model parameters. Longer follow-up in trials, and collection of HRQL scores, would help with reducing the uncertainty around long term cost-effectiveness of treatment.

In considering the cost-effectiveness results from studies which included a non-surgical comparator, it is evident that surgery after the failure of conservative care, could be cost-effective. However, there does not appear to be a consensus regarding at what stage surgical procedures might become cost-effective. This could be a potentially valuable research priority, alongside factors which influence the cost-effectiveness of surgery following these repeated failures of conservative care. It is also noted that research needs to explore the implications of using different health states in both conditions.

Health economists and modellers developing models in both conditions, also need to be more willing to explore the implications of extrapolation of treatment effect over an appropriate time horizon. The thesis by Mahon [40] provides a comprehensive review of the methods which can be used to infer parameters where they are unobserved. Additionally, guidance is available on how to capture the associated uncertainty relating to extrapolation of unobserved treatment parameters in sensitivity analyses. NICE methods guidance advocate scenario analyses with (i) nil treatment effect over the unobserved period; (ii) treatment effect during the unobserved period is set equal to the observed period; and (iii) treatment effect diminishes over time [51]

Future models should pay particular attention to the methodological challenges raised here to help ultimately enable more useful comparisons between treatments. But until modellers produce more high-quality modelling studies, consistent with modelling guidelines (e.g. [35] [38] [42] [44] [45]), the standard of discourse necessary to stimulate methodological improvements in these areas will be severely restricted.

Data Availability Statement

All data generated or analysed during this study are included in this published article.

Authors' contribution

JH, SJ, KK, RaO, ReO contributed to the study conception and design, design of search strategies and article selection, and interpretation of data. JH conducted search strategies as well as retrieved identified articles, performed data extraction, and drafted the manuscript. JH, SJ, KK, RaO, ML all contributed to manuscript preparation. All authors approved the final version of this article

References

1. Hartvigsen J, Hancock MJ, Kongsted A. et al. What low back pain is and why we need to pay attention. *Lancet.* 2015; 391(10137): 2356-2367

2. Woolf AD, Pfleger B. Burden of musculoskeletal conditions. *Bull World Health Organ* 2003; 81(9): 646-56

3. Parson S, Ingram M, Clarke-Cornwell AM, Symmons DPM. *A Heavy Burden The occurrence and impact of musculoskeletal conditions in the United Kingdom today.* Arthritis Research UK Epidemiology Unit, 2011, (www.escholar.manchester.ac.uk/uk-ac-man-scw:123774).

4. Bridges S. Chronic pain. In: Craig R, Mindell J, eds. *Health Survey for England 2011 Health, social care and lifestyles.* NHS Digital (https://files.digital.nhs.uk/publicationimport/pub09xxx/pub09300/hse2011-ch9-chronic-pain.pdf)

5. Foster NE, Anema, JR, Cherkin, D. et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018; 391(10137); 2368-2383

6. Koes BW, van Tulder M and Peul WC. Diagnosis and treatment of sciatica. *BMJ* 2007;334: 1313-1317,

7. Valat JP, Genevay S, Marty M. et al. Sciatica. *Best Pract Res Clin Rheumatol* 2010;24(2): 241-252).

8. Konstantinou K and Dunn KM. Sciatica: review of epidemiological studies and prevalence estimates. *Spine* 2008;33(22): 2464-2472.

9. van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. *Pain* 1995; 62(2): 233-240

10. NE Foster, K Konstantinou, M Lewis. et al. The clinical and cost-effectiveness of stratified care for patients with sciatica: the SCOPiC randomised controlled trial protocol (ISRCTN75449581). *BMC Musculoskelet Disord* 2017;18(1):172.

11. Lewis R, Wiliams N, Matar HE. et al. The clinical effectiveness and cost-effectiveness of management strategies for sciatica: systematic review and economic model. *Health Technol Assess*. 2011;15(39).

12. Sculpher M, Drummond M. Wither trial-based economic evaluation for health care decision making? *Health Econ*. 2006;15:677-687.

13. Andronis L, Kinghorn P, Qiao S. et al. Cost-Effectiveness of Non-Invasive and Non-Pharmacological Interventions for Low Back Pain: a Systematic Literature Review. *Appl Health Econ Health Policy*. 2017(15):173-201.

14. Kim N, Yang B, Lee T, Kwon S. An economic analysis of usual care and acupuncture collaborative treatment on chronic low back pain: a Markov model decision analysis. *BMC Complement Altern Med*. 2010;10(74).

15. Norton G, McDonough C, Cabral H, Shwartz M, Burgess JF. Cost-utility of cognitive behavioral therapy for low back pain from the commercial payer perspective. *Spine*. 2015;40(10):725–33.

16. Launois R, Henry B, Marty JR. et al. Chemonucleolysis versus surgical discectomy for sciatica secondary to lumbar disc herniation. A cost and quality-of-life evaluation. *Pharmacoeconomics*. 1994;6:453–63.

17. Lloyd A, Scott DA, Akehurst RL, Lurie-Luke E, Jessen G. Cost-effectiveness of low-level heat wrap therapy for low back pain. *Value Health*. 2004;7(4):413-422.

18. Wielage R, Bansal M, Wilson K. et al. The Cost-Effectiveness of Duloxetine in Chronic Low Back Pain: A US Private Payer Perspective. *Value Health*. 2013;16(2):334-344.

19. Wielage R, Bansal M, Wilson K, Klein R, Happich M. Cost-effectiveness of duloxetine in chronic low back pain: a quebec societal perspective. *Spine*. 2013;38(11):936-946.

20. Igarashi A, Akazawa M, Murata T. et al. Cost-effectiveness analysis of pregabalin for treatment of chronic low back pain in patients with accompanying lower limb pain (neuropathic component) in Japan. *Clinicoecon Outcomes Res*. 2015;7:505-520.

21. Fitzsimmons D, Phillips CJ, Bennett H. et al. Cost-effectiveness of different strategies to manage patients with sciatica. *PAIN*. 2014;155:1318-1327.

22. Udeh BL, Costandi S, Dalton JE. et al. The 2-year cost-effectiveness of 3 options to treat lumbar spinal stenosis patients. *Pain Pract.* 2015;15(2):107-116.

23. Skidmore G, Ackerman SJ, Bergin C. et al. Cost-effectiveness of the X-STOP (R) Interspinous Spacer for Lumbar Spinal Stenosis A Comparison with Conservative Care and Laminectomy. *Spine*. 2011;36(5):E345-E356.

24. Parker SL, Anderson LH, Nelson T, Patel VV. Cost-effectiveness of three treatment strategies for lumbar spinal stenosis: Conservative care, laminectomy, and the Superion interspinous spacer. *Int J Spine Surg*. 2015;9(28):1-11.

25. Koenig L, Dall TM, Gu Q, Saavoss J, Schafer MF. How Does Accounting for Worker Productivity Affect

the Measured Cost-Effectiveness of Lumbar Discectomy? *Clin Orthop Relat Res.* 2014;472(4):1069-1079.

26. Kuntz KM, Snider RK, Weinstein JN, Pope MH, Katz JN. Cost-effectiveness of fusion with and without instrumentation for patients with degenerative spondylolisthesis and spinal stenosis. *Spine*. 2000;25(9):1132-1139.

27. Kim S, Hedjri S, Coyte PC, YR. R. Cost-utility of lumbar decompression with or without fusion for patients with symptomatic degenerative lumbar spondylolisthesis. *Spine*. 2012;12(1):44-54.

28. Schmier J, Halevi M, Maislin G, Ong K. Comparative cost effectiveness of Coflex interlaminar stabilization versus instrumented posterolateral lumbar fusion for the treatment of lumbar spinal stenosis and spondylolisthesis. *Clinicoecon Outcomes Res*. 2014;6(1):125-131.

29. Yaghoubi M, Moradi-Lakeh M, Moradi-Joo M. et al. The cost-effectiveness of dynamic and static interspinous spacer for lumbar spinal stenosis compared with laminectomy. *Med J Islam Repub Iran*. 2016;30:339.

30. Bydon M, Macki M, Abt NB. The cost-effectiveness of interbody fusions versus posterolateral fusions in 137 patients with lumbar spondylolisthesis. *Spine*. 2015;15(3):492-498.

31. Parkinson B, Goodall S, Thavaneswaran P. Cost-effectiveness of lumbar artificial intervertebral disc replacement: driven by the choice of comparator. *ANZ J Surg*. 2013;83(9):669-675.

32. Vertuani S, Nilsson J, Borgman B. et al. A Cost-Effectiveness Analysis of Minimally Invasive versus Open Surgery Techniques for Lumbar Spinal Fusion in Italy and the United Kingdom. *Value Health*. 2015;18(6):810-816.

33. Malter AD, Larson EB, Urban N, Deyo RA. Cost-effectiveness of lumbar discectomy for the treatment of herniated intervertebral disc. *Spine*. 1996;21:1048-1055.

34. Fryback DG, Dasbach EJ, Klein R. et al. The Beaver Dam Health Outcomes study: Initial Catalog of Health-state Quality Factors. *Med Decis Making*. 1993;13(2):89-102.

35. Phillips Z, Ginnelly L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment. *Pharmacoeconomics*. 2006;24:355-371.

36. Jaime-Caro J, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices-Overview: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health*. 2012;15:796-803.

37. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ*. 2003;12(11):837-848.

38. Siebert U, Alagoz O, Bayoumi AM. et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. *Value Health*. 2012;15:812-820.

39. Dunn KM, Campbell P, Jordan KP. Long-term trajectories of back pain: cohort study with 7-year follow-up. *BMJ Open*. 2013;3.

40. Mahon R. *Temporal uncertainty in cost-effectiveness decision models.* PhD thesis. 2014.

41. Lofland JH, Pizzi LFKD. A Review of Health-Related Workplace Productivity Loss Instruments. *Pharmacoeconomics*. 2004;22(3):165-184.

42. Soto J. Health economic evaluations using decision analytic modeling: principles and practices. Utilization of a checklist to their development and appraisal. *Int J Technol Assess*. 2002;18(1):94-111.

43. Weinstein JN, Lurie JD, Tosteson TD. et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am*. 2009;91:1295-1304.

44. Roberts M, Russell LB, Paltiel D. et al. Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. *Med Decis Making*. 2012;32(5):678-689.

45. Husereau D, Drummond M, Petrou S al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ* 2013; 346 :f1049

46. Institute for Clinical and Economic Review. *Management options for patients with low back disorders.* 2011. Available from: https://icer-review.org/wp-content/uploads/2016/02/LBP-Final-Appraisal-6-24-11.pdf

47. van den Hout WB, Peul WC, Koes BWet al. The Hague Spine Intervention Prognostic Study Group. Prolonged conservative care versus early surgery in patients with sciatica from lumbar disc herniation: cost-utility analysis alongside a randomised controlled trial. *BMJ* 2008; 336: 1351-1354.

48. Whynes DK, McCahon RA, Ravenscroft A, Hardman J. Cost effectiveness of epidural steroid injections to manage chronic lower back pain. *BMC Anesthesiol*. 2012;12:26–32.

49. Stromqvist B, Fritzell P, Hagg O, Jonsson B. National Swedish Register for Lumbar Spine Surgery report 2008. Available from: http://www.4s.nu/pdf/Ryggregisterrapport_2008_eng_version.pdf.

50. Tapp, SJ, Martin, BI, Tosteson TD, et al. Understanding the value of minimally invasive procedures for the treatment of lumbar spinal stenosis: the case of interspinous spacer devices. *Spine J*. 2018:18:584-592

51. NICE 2013. *National Institute for Health and Care Excellence. Guide to the methods of technology*

appraisal 2013. Available from: https://www.nice.org.uk/process/pmg9/chapter/foreword

52. Becker A , Held H , Redaelli M , et al. Low back pain in primary care. Costs of care and prediction of future health care utilization .*Spine* 2010;35:1714 – 20 .

53. Work Loss Data Institute . Official Disability Guidelines . 15th ed .Encinitas, CA : Work Loss Data Institute ; 2010

54. Korean National Health & Nutrition survey. 2007 Available at [https://knhanes.cdc.go.kr/knhanes/main.do]

55. . Ministry of Employment and Labor. Available at [http://laborstat.molab.go.kr].

56. . Centers for Disease Control and Prevention. National Health Interview Survey. Available at: http://www.cdc.gov/nchs/nhis. htm.

57. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. *JAMA.* 2006; 296: 2451–2459.

58. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics*. 2004;22(3): 165–184

59. Nadler SF, Steiner DJ, Erasala GN, et al. Continuous low-level heat wrap therapy provides more efficacy than ibuprofen and acetaminophen for acute low back pain. *Spine* 2002;27:1012–7

60. Grotle M, Brox JI, Veierod MB, et al. Clinical course and prognostic factors in acute low back pain: patients consulting primary care for the first time. *Spine*; 2005:30(8):976-982

61. Cassidy JD, Cote P, Carroll LJ, Kristman V: Incidence and course of low back pain episodes in the general population. *Spine;* 2005:30(24):2817-2823.

62. Witt CM, Jena S, Selim D, et al. Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *Am J Epidemiol;* 2006, 487-496.

63. Thomas KJ, MacPherson H, Thorpe L, et al. Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. *BMJ* 2006, 623

64. IMS-Brogan Inc . Provincial Formulary Database . Kirkland, Quebec : Delta PATM ; 2011 .

65. Ontario Case Costing Initiative . OCCI Costing Analysis Tool .2010 . Available at: http://www.occp.com.

Accessed August 24,

66. Lamb SE, Lall R, Hansen Z, et al. A multicentred randomised controlled trial of a primary carebased cognitive behavioural programme for low back pain. The Back Skills Training (BeST) trial. *Health Technol Assess.* 2010; 14:41.
67. Zucherman JF, Hsu KY, Hartjen CA, et al. A multicenter, prospective, randomized trial evaluating the X-STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication. *Spine* 2005;30(12):1351–8.
68. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2008;33:2789–2800.
69. Tosteson AN, Skinner JS, Tosteson TD, et al. The cost effectiveness of surgical versus nonoperative treatment for lumbar disc herniation over two years: evidence from the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2008;33:2108-2115.
70. Centers for Medicare & Medicaid Services. MEDPAR Limited Data Set (LDS) - Hospital (National). Available at: http://cms. gov/Research-Statistics-Data-and-Systems/Files-for-Order/Limited DataSets/MEDPARLDSHospitalNational.html. Authors accessed January 29, 2013.
71. Whynes DK, McCahon RA, Ravenscroft A, Hardman J. Cost effectiveness of epidural steroid injections to manage chronic lower back pain. *BMC Anesthesiol*. 2012;12:26–32.
72. Glassman SD, Polly DW, Dimar JR, Carreon LY. The cost effectiveness of single-level instrumented posterolateral lumbar fusion at 5 years after surgery. *Spine* 2012;37: 769–774.
73. Tosteson AN, Tosteson TD, Lurie JD, et al. Comparative effectiveness evidence from the spine patient outcomes research trial surgical versus nonoperative care for spinal stenosis, degenerative spondylolisthesis, and intervertebral disc herniation. *Spine*. 2011;36:2061–2068
74. Tosteson AN, Lurie JD, Tosteson TD, et al. Surgical treatment of spinal stenosis with and without degenerative spondylolisthesis: cost-effectiveness after 2 years. *Ann Intern Med*. 2008;149:845–853.
75. Taguchi T, Igarashi A, Watt S, et al. Effectiveness of pregabalin compared with usual care for the treatment of chronic low back pain with a neuropathic component in Japan; the patient's perspective. *J Pain Res*. 2015;8:487–497.
76. Parker SL, Godil SS, Mendenhall SK, et al. Two-year comprehensive medical management of degenerative lumbar spine disease (lumbar spondylolisthesis, stenosis, or disc herniation): a value analysis of cost, pain, disability, and quality of life.J. Neurosurg. *Spine* 2014.
77. Patel VV, Whang PG, Haley TR, et al. Two-year clinical outcomes of a multicenter randomized controlled trial comparing two interspinous spacers for treatment of moderate lumbar spinal stenosis.*BMC Musculoskelet Disord* 2014 (15): 221.
78. Yano S, Hida K, Seki T et al. A new ceramic interspinous process spacer for lumbar spinal canal stenosis *Neurosurgery* 2008 61(1): ONS108-ONS113
79. Katz JN, Lipson SJ, Lew RA, et al. Lumbar laminectomy alone or with instrumented or noninstrumented

arthrodesis in degenerative lumbar spinal stenosis: Patient selection, costs, and surgical outcomes. *Spine* 1997;22:1123–31.

80. Deyo RA, Mirza SK, Martin BI, et al. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 2010;303:1259–65.
81. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007;356:2257–70.
82. Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacementcompared to lumbar fusion: a randomised controlled trial with 2-yearfollow-up.Eur. *Spine* .2009;18: 1512–9
83. Australian Department of Health and Ageing.Medicare Benefits Sched-ule Book, operating from 1 May 2010. 2010, DoHA: Canberra
84. Davis RJ, Errico TJ, Bae H, Auerbach JD. Decompression and Coflex interlaminar stabilization compared with decompression and instrumented spinal fusion for spinal stenosis and low-grade degenerative spondylolisthesis: two-year results from the prospective, randomized, multicenter, Food and Drug Administration Investigational Device Exemption trial. *Spine*. 2013;38(18):1529–1539.
85. Stromqvist B, Fritzell P, Hagg O, Jonsson B. *National Swedish Register for Lumbar Spine Surgery report 2008.* 2008. Available from: http:// www.4s.nu/pdf/Ryggregisterrapport_2008_eng_version.pdf. [Authors Accessed July 15, 2014].
86. Remmers A, Sharma U, LaMoreaux L, et al. *Pregabalin treatment of patients with chronic low back pain* [Poster #660]. Presented at: 19th Annual Meeting of the American Pain Society. Atlanta, GA, 2000.